Revista Brasileira de Geomática
ISSN: 2317-4285

https://periodicos.utfpr.edu.br/rbgeo

Point cloud filtering and device performance in SfM: comparing cameras and smartphones for 3D modeling

Francisco Roza de Moraes
maverick@usp.br
orcid.org/ 0000-0002-7413-6467
Universidade de São Paulo (USP) – Escola de Engenharia de São Carlos, São Carlos, São Paulo, Brasil.

Irineu da Silva
irineu@sc.usp.br
orcid.org//0000-0001-5775-6683
Universidade de São Paulo (USP) – Escola de Engenharia de São Carlos, São Carlos, São Paulo, Brasil.

ABSTRACT

Structure from Motion (SfM) is widely used for 3D modeling from digital images, offering accessibility and compatibility with common devices like cameras and smartphones. However, the accuracy of the generated models significantly depends on the quality of the image set and the refinement processes, such as point cloud filtering, which are often overlooked in literature. This study evaluates the effectiveness of different capture devices for SfM-based 3D modeling and investigates the impact of sparse point cloud filtering on model accuracy. Models were created from image sets obtained using both a camera and a smartphone, with and without filtering. The results indicate that filtering is essential for achieving high-quality models from cameras, providing an RMSE of 0.1 mm and enhanced object detail. However, models derived from smartphone images showed competitive potential. These findings highlight the importance of refinement strategies in SfM-based modeling and contribute to optimizing its use across various capture contexts.

KEYWORDS: Structure from Motion. Reconstruction Uncertainty. Projection Accuracy. Reprojection Error. Smartphone.

INTRODUCTION

The Structure from Motion (SfM) Multi-View Stereo (MVS) technique has established itself as a versatile and accessible tool for creating three-dimensional models from sets of digital images. Its growing popularity is largely attributed to its flexibility in utilizing widely available capture devices, such as cameras with interchangeable lenses and smartphones, in contrast to traditional photogrammetry techniques that rely on metric cameras designed with high technical precision and elevated costs (Westoby et al., 2012). This accessibility, combined with the development of intuitive software, enables professionals and researchers from diverse fields to explore the technique's potential in both indoor and outdoor environments, ranging from close-range captures (An et al., 2021; Liu et al., 2016) to long-distance applications (Garcia; Oliveira, 2021; Woodget; Austrums, 2017).

SfM involves the estimation of camera poses and the reconstruction of sparse three-dimensional point clouds through image matching and optimization algorithms. A critical component of this process is Bundle Adjustment, which optimizes both intrinsic and extrinsic camera parameters, along with the 3D coordinates of the sparse points, by minimizing the reprojection error across all images (Triggs et al., 2000). While this refinement enhances the internal consistency of the reconstruction, it does not inherently eliminate outliers or inconsistencies that may persist in the matched feature correspondences.

After SfM, the Multi-View Stereo (MVS) stage generates a dense point cloud by leveraging the estimated camera parameters. However, this phase can propagate and amplify errors introduced in prior stages, particularly when the input imagery contains regions with low texture, repetitive patterns, or insufficient inter-image overlap (Luhmann et al., 2023; Westoby et al., 2012).

Given these challenges, post-alignment filtering techniques play a pivotal role in refining the resulting point cloud, which often retains noise, redundancy, and geometric artifacts not addressed during initial alignment or sparse optimization Di Filippo et al. (2022). These filtering methods enhance the geometric accuracy and visual quality of the final model by eliminating outliers and enforcing surface coherence—a critical requirement in applications demanding high precision.

While these computational processes shape the structure of the model, the quality of the final output is also highly dependent on the image acquisition stage. Despite the ease of use provided by SfM, producing high-quality three-dimensional models requires meticulous attention to the technical aspects of the process. As highlighted by Creus, Sanislav, and Dirks (2021) and Moraes and Silva(2024), the use of precise reference points is critical for image alignment and the proper pre-calibration of photographic equipment. Furthermore, the quality of the generated models is directly linked to the characteristics of the image set used, which must exhibit a high level of detail, adequate coverage of the modeled object or scene, and sufficient sharpness to ensure the accuracy of reconstruction algorithms (Micheletti et al., 2015).

The consumer-grade capture devices, such as smartphones or cameras with interchangeable lenses, can streamline the image acquisition process and produce high-quality three-dimensional products (Jaud et al., 2019; Verma; Bourke, 2019). However, due to their sensors and optical systems limitations, these devices may

introduce inconsistencies in the captured images, often resulting in noisy and imprecise point clouds (Tavani et al., 2020; Woodget; Austrums, 2017). These issues are intensified by the detection and correlation algorithms, which may perform unreliable matches between elements of images, thereby negatively affecting the quality and accuracy of the resulting three-dimensional models.

In practice, unfiltered point clouds often contain artifacts, including isolated points, duplicate surfaces, and structural noise, which can distort measurements and hinder accurate geometric interpretation. Nota, Nijland, and de Haas (2022) demonstrated that insufficient filtering may introduce geometric deviations exceeding 5 mm in short-range applications—errors that could critically affect the reliability of structural analyses or deformation monitoring.

Point cloud filtering techniques are recommended during the SfM processing workflow to address these challenges, as discussed in Di Filippo et al. (2022) and Nota, Nijland, and de Haas (2022). However, users often overlook these approaches, leading to reduced accuracy and quality in the resulting models. The development of a rigorous methodology incorporating filtering metrics, such as those proposed by the United States Geological Survey (Over et al., 2021), is essential for creating more accurate and representative three-dimensional models.

The integration of point cloud filtering processes, combined with careful consideration of the technical factors involved in SfM modeling, can significantly enhance the quality of the resulting three-dimensional products. This improvement can be particularly critical for enabling more precise evaluations and comparisons of the characteristics of widely used SfM capture devices, such as smartphones and cameras with interchangeable lenses.

Accordingly, this study aims to evaluate the influence of point cloud filtering on the quality parameters of three-dimensional models generated using the SfM technique, focusing on image sets captured by common photographic equipment (Camera and Smartphone) in short-range scenarios, such as those used in structural laboratory testing. Additionally, it seeks to determine the most suitable technology for such applications by comparing the products generated by different devices. The goal is to identify the strengths and limitations of the produced models, providing insights to guide the selection of the most appropriate technology given current technological constraints.

METHOD

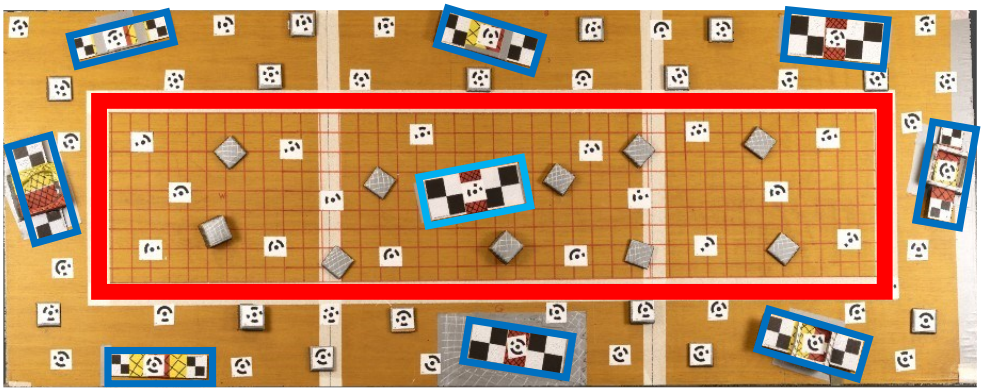
The primary objective of this research is to evaluate the differences and quality levels achieved in 3D modeling using the SfM technique, comparing the performance of an interchangeable lens camera with that of a smartphone as photographic capture equipment. This study focuses on assessing the final quality of the 3D models rather than examining the workflow or specific software configurations. All reconstructions were processed using Agisoft Metashape Pro, a well-established software that facilitates efficient point cloud filtering and model generation. For detailed information on SfM workflows and configurations, readers are encouraged to consult (James et al., 2017; Leon et al., 2015; Tinkham; Swayze, 2021).

DELIMITATION OF THE REGION OF INTEREST AND EXPERIMENTAL SETUP

The photographic captures were conducted in the Department of Transportation Engineering laboratory at the São Carlos School of Engineering, University of São Paulo. The environment featured both natural and artificial lighting. To ensure uniform illumination and avoid shadows on the modeled object's surface, auxiliary lighting systems (softboxes) equipped with 7,000-lumen LED lamps and a color temperature of 5,000 Kelvin were used. These were strategically positioned to minimize shadows and light variations.

A wooden board measuring 210 cm x 80 cm x 4 cm was employed to simulate an experiment in structural laboratory settings. This material was selected for its high surface texture variability, which aids in detecting corresponding elements between images during the SfM modeling process. **Figure 1** shows the modeled object and the additional elements on its surface that were used to assist processing.

Figure 1 – Assessment of the Reprojection Error of Sparse Point Cloud



Source: Author (2024).

The region of interest (ROI) was defined by using white adhesive tape and highlighted in the figure with a red rectangle at the center of the object. A grid pattern, drawn with a red permanent marker, was applied within this area to introduce artificial texture and enhance the detection of corresponding features during the SfM modeling process, as discussed in Hafeez et al. (2018). To increase the complexity of the modeled scene, cubes of varying heights and positions were distributed across the surface of the wooden board. These cubes, covered with silver tape, simulated diverse elements present in structural environments.

Eight sets of acrylic rulers, represented in dark blue in the figure, were randomly placed around the ROI and aligned with the axes of the board. These rulers, featuring chessboard patterns and known dimensions, served as Scale Bars (SBs) for the calibration and metric scaling of the 3D models. An additional acrylic ruler, positioned at the center of the ROI and depicted in light blue, acted as a Control Bar (CB), providing a reference for verifying the model's dimensional accuracy.

DATA ACQUISITION

To generate high-quality 3D models for comparative analysis of the performance of cameras and smartphones in short-range laboratory conditions, a systematic photographic acquisition method was employed. The imaging process followed a vertical grid pattern, as illustrated in **Figure 2**, and was replicated for both capture devices.

Figure 2 – Photographic capture process, with the squares representing the positioning of the capture equipment

Source: Author (2024).

The photographic process, considered a crucial step for ensuring the quality of SfM-generated models (Caldera-Cordero; Polo, 2019), was based on the methodology described by Moraes and Silva (Moraes; Silva, 2024). Images were captured at 1 meter, with an overlap of approximately 80%, pre-calibrated devices, and combined vertical and oblique images taken at ±15° yaw (relative to the Y-axis).

For the conventional equipment, a Canon EOS R camera with a full-frame sensor (36 mm × 24 mm) and a 35 mm focal length lens was utilized. The resulting images had dimensions of 6,472 × 4,498 pixels, yielding a Ground Sampling Distance (GSD) of nominally 0.15 mm at the defined capture distance.

The smartphone used for comparison was an Apple iPhone 15 Pro Max, equipped with a 9.8 mm × 7.3 mm sensor. The focal length of the smartphone was 6.7 mm (equivalent to a 24 mm lens on a full-frame sensor), as recorded in the image metadata, and the image resolution was 5,712 × 4,284 pixels. This configuration resulted in a GSD of approximately 0.25 mm at the same capture distance.

To ensure consistent image sharpness, both devices were mounted on a tripod, and operated using a 5-second timer to minimize motion blur. To eliminate inconsistencies from automatic adjustments, all photographic parameters were manually configured and fixed to the standard settings (Manual focus and exposure D+1). Additionally, images were recorded in RAW format—CR3 for the digital camera and DNG for the smartphone—to preserve the original sensor data without any in-camera post-processing. The use of RAW formats enables the retention of unaltered radiometric and geometric information, which is essential for ensuring data integrity in photogrammetric applications.

As part of the pre-calibration procedure, the internal parameters of each device were estimated using Agisoft Metashape software, which adopts an eight-

parameter model based on the formulation originally proposed by Brown (Brown, 1971) This model includes the focal length (F), the principal point offsets (Cx, Cy), three radial distortion coefficients (K1, K2, K3), and two tangential distortion coefficients (P1, P2). **Table 1** and **Table 2** presents the estimated calibration parameters and their respective errors for the digital camera and the smartphone, respectively.

Table 1 – Estimated calibration parameters and correlation matrix (partial) for the devices used Canon EOS R camera | (b) iPhone 15 Pro Max

ID	Value	Error	F	Cx	Cy	K1	K2	K3	P1	P2
F	6637.84	0.1530	1.00	0.05	-0.07	-0.42	0.43	-0.42	0.04	-0.04
Cx	66.6299	0.2801		1.00	0.00	-0.07	0.08	-0.09	0.96	0.00
Cy	19.5779	0.2364			1.00	-0.04	0.04	-0.04	0.00	0.94
K1	-0.00353	0.0002				1.00	-0.99	0.98	-0.04	-0.04
K2	0.14094	0.0010					1.00	-0.99	0.05	0.03
K3	-0.03760	0.0015						1.00	-0.05	-0.03
P1	0.00076	1.77E-05							1.00	0.00
P2	0.00133	1.29E-05								1.00

Table 2 – Estimated calibration parameters and correlation matrix (partial) for the devices used iPhone 15 Pro Max

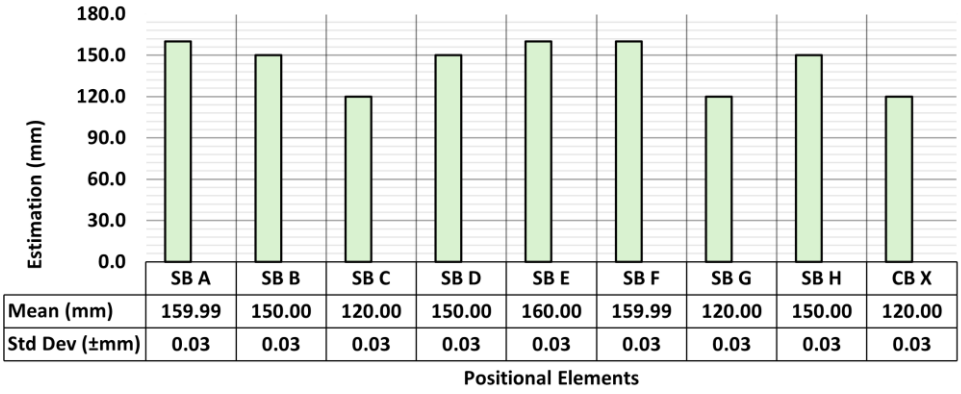
ID	Value	Error	F	Cx	Cy	K1	K2	K3	P1	P2
F	5658.09	0.9864	1.00	0.07	-0.02	-0.53	0.51	-0.59	0.06	0.03
Cx	5.7313	1.2127		1.00	0.01	-0.03	0.03	-0.03	0.89	0.01
Cy	17.8548	1.1178			1.00	-0.11	0.10	-0.09	0.00	0.91
K1	0.23235	0.0019				1.00	-0.99	0.97	-0.04	-0.11
K2	-0.84326	0.0077					1.00	-0.99	0.05	0.10
K3	0.93077	0.0105						1.00	-0.05	-0.10
P1	-0.00016	4.86E-05							1.00	0.00
P2	-0.00028	5.09E-05								1.00

In addition to the parameter values, the correlation matrices were also analyzed to assess the interdependence between estimated coefficients. As noted by Remondino et al. (2006), strong correlations among the radial distortion coefficients (K1–K3) are inherent to the structure of the distortion model and therefore expected. Likewise, tangential distortion coefficients (P1, P2) often exhibit coupling with the principal point offsets (Cx, Cy), particularly in systems with shorter focal lengths or limited convergence among image axes. Nevertheless, these correlations are generally deemed acceptable and do not compromise the overall reliability of the camera calibration.

The dimensions of the scale elements (SBs and CBs) were measured using a Starrett EC799A-8 digital caliper. This device provides a measurement accuracy of ±0.02 mm for lengths up to 10 cm and ±0.03 mm for greater dimensions (Starrett Company, 2024). These precise measurements, combined with the photographic capture process, played a vital role in determining the quality of the resulting 3D models.

To minimize potential errors, each acrylic ruler was measured five times. The processed data, including the mean lengths and standard deviations of the scale elements, are presented in **Figure 3**.

Figure 3 – Mean values and Standard Deviations of the scale elements



	SB A	SB B	SB C	SB D	SB E	SB F	SB G	SB H	CB X
Mean (mm)	159.99	150.00	120.00	150.00	160.00	159.99	120.00	150.00	120.00
Std Dev (±mm)	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03

Source: Author (2024).

POINT CLOUD FILTERING

As previously stated, point cloud filtering is a critical step in the 3D modeling process, particularly for datasets generated through SfM and digital photogrammetry. While these techniques produce dense, information-rich point clouds, they often contain uncertainties caused by the capture conditions, equipment calibration inaccuracies, or algorithmic limitations. The filtering process aims to enhance positional accuracy and ensure that the point cloud faithfully represents the reconstructed object or scene.

This study employed the methodology established by the United States Geological Survey (USGS) (Over et al., 2021) for point cloud filtering. The filtering process was conducted using Agisoft Metashape, with parameter values selected based on prior experimental validation and USGS guidelines. Quality metrics, including **Reconstruction Uncertainty**, **Projection Accuracy**, and **Reprojection Error**, were applied to refine the point clouds, enhancing the precision of the resulting 3D models.

Reconstruction Uncertainty: This metric can be interpreted as the ratio between the largest and smallest semiaxes of the error ellipse created during the triangulation of 3D point coordinates from multiple images. A high ratio suggests a high uncertainty, typically along the Z-axis (depth), and these points may introduce local noise without contributing meaningful geometric information (Di Filippo et al., 2022). For this study, points with a reconstruction uncertainty ratio above a **threshold of 10** were removed. This threshold is considered effective in eliminating noisy points without changing the overall model accuracy.

Projection Accuracy: This metric reflects how accurately a 3D point is in the images used for its reconstruction. It is calculated as the ratio of the sum of the image scales (resolutions) in which the point is visible to the total number of images where the point appears (Stark et al., 2022). Although USGS method (Over et al., 2021) suggest a threshold value of 2, our previous tests showed that this value led to excessive point removal, which negatively impacted subsequent

processing stages. Therefore, a **threshold of 3** was adopted in this study, which offered a better balance between accuracy and point cloud density.

Reprojection Error: This metric measures the discrepancy between a point's estimated position in 3D space and its projected location on the original images. It serves as an indicator of both model accuracy and camera alignment. In line with commonly accepted standards, including those followed in multiple studies using Agisoft Metashape, a maximum threshold of 0.3 pixels was adopted. Values beyond this may indicate alignment errors or inconsistencies in the reconstruction, but additional optimization below this threshold typically does not yield perceptible improvements in model quality (Barba et al., 2019).

When applied collectively, these filtering processes, as demonstrated by Capolupo (2021) and Over et al. (2021), ensure that the point cloud is optimized for 3D modeling, achieving greater reliability and precision.

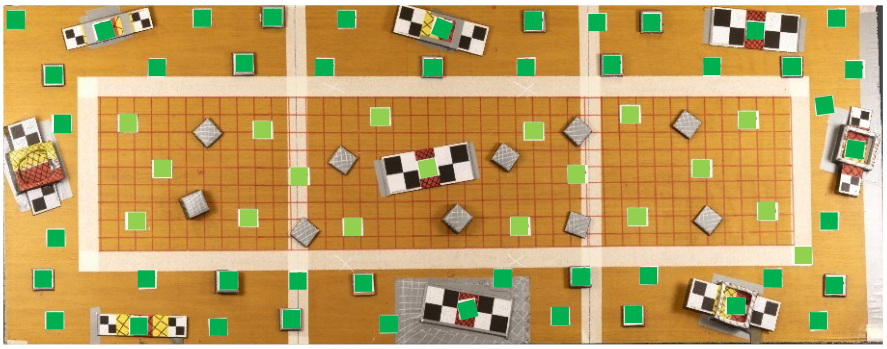
The parameter thresholds adopted in this study are the result of prior experimental assessments and are consistent with the quality of the photographic data acquired. Overall, these values are expected to be appropriate for applications with comparable goals, especially when employing conventional imaging equipment of good quality.

COORDINATE TRANSFORMATION

Due to the absence of a referenced coordinate system, stemming from the challenges of establishing the same scale bar precision, each 3D model was initially referenced to the coordinate system of the capturing device. To enable comparison among the 3D products generated, a transformation of coordinates into a common reference system was necessary.

To facilitate this transformation, 46 auto-detectable targets, designated as Control Points (CPs), were distributed around the ROI, as represented by dark green squares in **Figure 4**. Additionally, 18 Verify Points (VPs), depicted as light green squares within the ROI, were used to evaluate model quality and accuracy after the transformation.

Figure 4 – Arrangement of automatic detection targets for coordinate transformation



Source: Author (2024).

The coordinates of the CPs and VPs, initially in distinct frames, were transformed into a common reference system using a Python-based affine transformation routine combined with the Least Squares method. This approach

allowed minimizing errors and ensuring compatibility of coordinates across different models.

The affine transformation preserves proportions and parallelism, while the Least Squares method optimizes accuracy by minimizing discrepancies (Andrei, 2006). This methodology supports precise comparative analysis of the 3D models by effectively aligning coordinates between captures.

Experimental Models for Coordinate Transformation

To determine the optimal number and arrangement of CPs for the coordinate transformation, four experimental configurations were tested, as summarized in **Table 3**. The CPs were independently selected from the available targets to ensure unbiased evaluation and mitigate overfitting in the transformation process. The Verify Points (VPs), totaling 18 in all configurations, were exclusively positioned within the central region of interest of the imaged object, enabling consistent and rigorous validation of transformation accuracy.

Table 3 – Arrangement of the layout and quantity of Control Points for the Coordinate Transformation process

ID	Model	Qty. Control Points	Qty. Verify Points
AT	Alternating Targets	23	18
IT	Internal Targets	25	18
AIT	Alternating Internal Targets	11	18
ECT	Edges and Central Targets	6	18

The first configuration, referred to as Alternating Targets (AT), used 23 CPs distributed alternately across all available targets. The second, named Internal Targets (IT), relied on 25 CPs positioned around the region of interest. In the third configuration, Alternating Internal Targets (AIT), 11 CPs were alternated around the region of interest to create a more restrictive setup. Finally, the Edges and Central Targets (ECT) configuration employed six CPs strategically placed along the edges and at the center of the horizontal boundaries of the region of interest.

Each configuration underwent testing to assess positional quality using affine transformations, emphasizing precision and alignment fidelity. These experiments provided insights into the performance of different setups and informed the selection of the most robust approach to ensure repeatable and high-quality results in 3D modeling.

AVAILABILITY OF DATA AND COMPUTATIONAL RESOURCES

All image datasets captured by the different devices, along with measurement values and the computational codes used to process the reference elements in this study, are available in the repository Zenodo (10.5281/zenodo.14862180).

RESULTS AND DISCUSSION

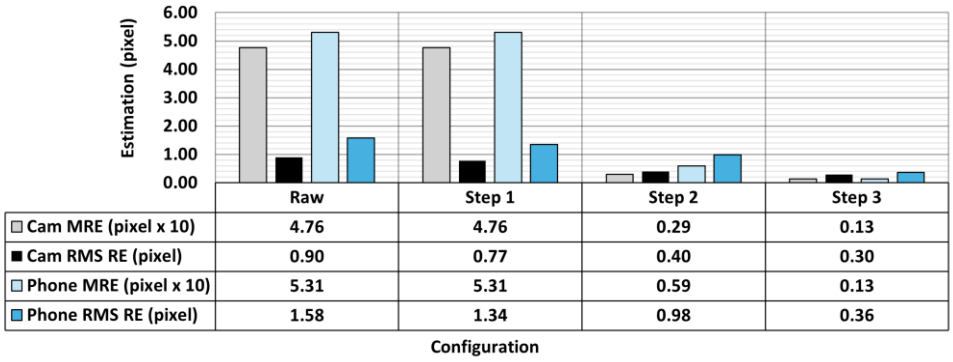
POINT CLOUD FILTERING

The alignment and preparation of sparse point clouds utilized image sets with an approximate overlap of 80% to ensure robust feature matching across the dataset. Camera and smartphone calibration were conducted using a pre-calibration model that incorporated all eight scale bars, which provided consistent accuracy. Ambient lighting conditions were maintained constant throughout the image capture process to minimize illumination variations and achieve uniform photometric quality across the dataset.

For this study, SfM processing generated raw sparse point clouds of approximately 1.45 million tie points for the camera dataset and 1.96 million tie points for the smartphone dataset. Considering the physical dimensions of the imaged area, this corresponds to an average tie point density of approximately 86.3 points/cm² for the camera and 116.7 points/cm² for the smartphone. Densely distributed tie points contribute to greater geometric redundancy and internal consistency in the photogrammetric network, which are essential for achieving accurate and robust 3D reconstructions, especially in small-scale, high-resolution applications (Luhmann et al., 2023).

Figure 5 presents the Reprojection Error values, measured in pixels, corresponding to each stage of the sparse point cloud filtering process for the image sets acquired by both the camera and smartphone.

Figure 5 – Assessment of the Reprojection Error of Sparse Point Cloud



	Raw	Step 1	Step 2	Step 3
Cam MRE (pixel x 10)	4.76	4.76	0.29	0.13
Cam RMS RE (pixel)	0.90	0.77	0.40	0.30
Phone MRE (pixel x 10)	5.31	5.31	0.59	0.13
Phone RMS RE (pixel)	1.58	1.34	0.98	0.36

Source: Author (2024).

The unfiltered Raw configuration produced SfM-generated point clouds with elevated Maximum Reprojection Error (MRE) values—47 pixels for the cloud produced with the camera set and 53 pixels for the cloud from the smartphone set. These high error values can adversely affect the geometric quality of 3D models, especially in low-texture regions, resulting in distortions and inaccuracies in surface representation, as discussed in Liao and Wood (2020). The values of Root Mean Square of the Reprojection Error (RMS RE) for the unfiltered model were approximately 0.9 pixels for the camera-captured dataset and 1.58 pixels for the smartphone dataset. While these values are generally considered low, they can still impact the geometric accuracy of the final model, especially in regions of high

geometric complexity where precise reprojection is crucial for accurate reconstruction.

The initial step of the filtering process, referred to as Step 1, analyzes the Reconstruction Uncertainty parameter. Despite the lack of significant variations in the MRE and RMS RE values when compared to the raw point clouds for both image sets, the filtering based on the Reconstruction Uncertainty parameter was crucial for enhancing geometric accuracy in subsequent filtering stages. Following this step, the point clouds were reduced to 354,845 points for the image set from the camera (21.1 points/cm²) and 456,544 points for the image set from the smartphone (27.1 points/cm²).

The subsequent stage of the filtering process, referred to as Step 2, applies to the Projection Accuracy parameter. The observed values for MRE and RMS RE for both image sets showed significant reductions compared to the previous step. Specifically, the camera-derived image set achieved an MRE of 0.29 pixels and an RMS RE of 0.40 pixels, while the smartphone-derived model attained an MRE of 0.59 pixels and an RMS RE of 0.98 pixels. After this step, the point clouds were reduced to 229,330 points for the camera image set (13.7 points/cm²) and 285,881 points for the smartphone image set (17.0 points/cm²).

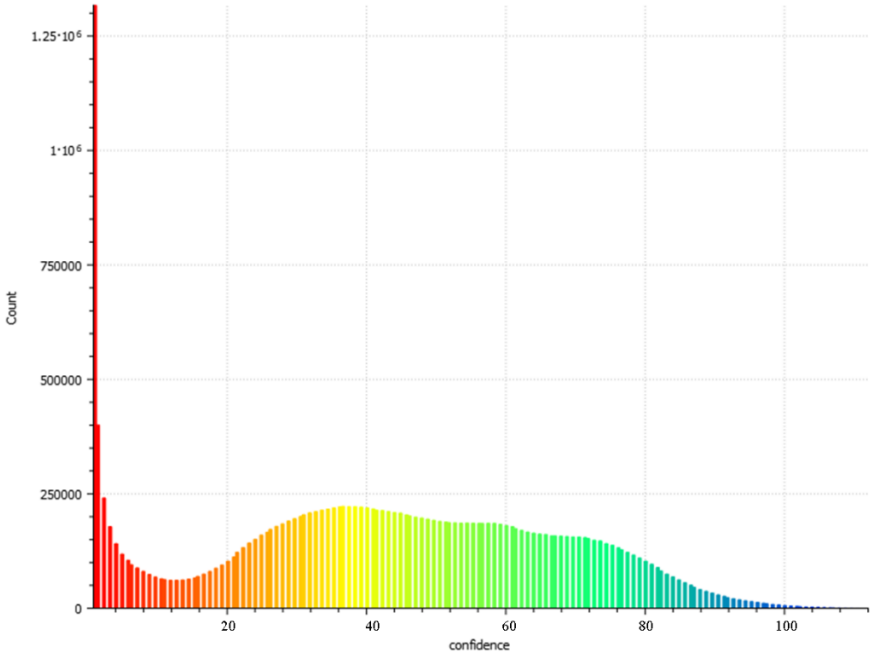
The final stage of the filtering process, Step 3, applies the Reprojection Error parameter. The observed values showed a slight improvement over the previous step for the point cloud derived from the camera image set, achieving an MRE of 0.13 pixels and an RMS RE of 0.30 pixels. For the point cloud produced from the smartphone image set, the values demonstrated a more substantial improvement over the previous step, with an MRE of 0.13 pixels and an RMS RE of 0.36 pixels.

After the filtering process, the point cloud derived from the camera image set contained 146,244 points (8.7 points/cm²), retaining approximately 10% of the initial points detected, indicating high filtering selectivity while preserving critical structure. In contrast, the point cloud generated from the smartphone image set was reduced to 69,020 points (4.1 points/cm²), or about 3.5% of the original detections, highlighting an even more substantial filtration due to differences in initial data quality and sensor precision. The results demonstrate that the camera produced a significantly higher density of high-quality points post-filtering compared to the smartphone, reflecting its superior image consistency and sensor performance.

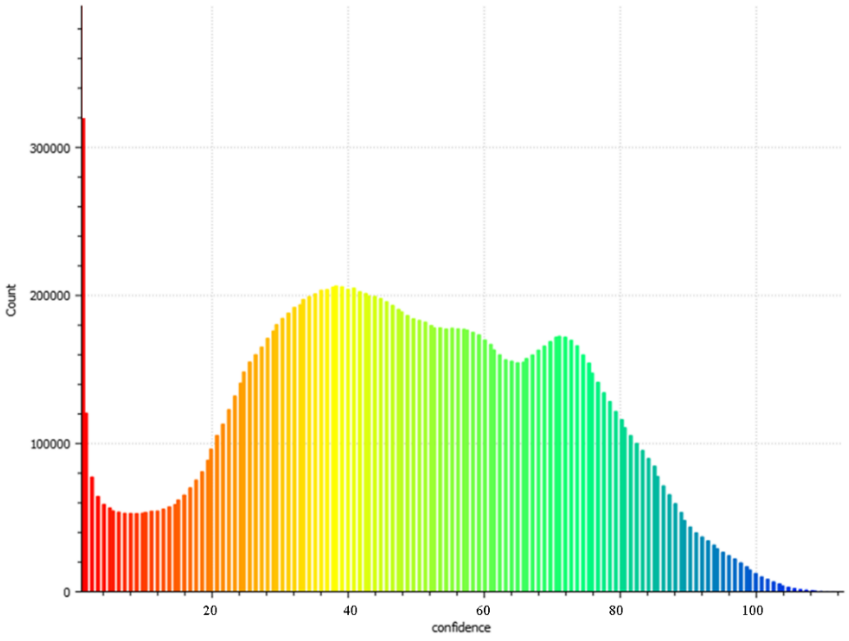
Figure 6 illustrates histograms of the Confidence parameter values associated with the dense point cloud derived from the camera image set, before and after the filtering process, respectively.

The first histogram (**Figure 6a**) highlights a large concentration of points with low confidence values. In contrast, the second histogram (**Figure 6b**) shows a redistribution of values, with a pronounced shift toward higher confidence levels—an outcome of the filtering process that effectively removed low-quality observations. It is worth noting that the y-axes of both histograms are not standardized: Figure 5a reaches up to 1.25×10^6 points, whereas Figure 5b peaks at approximately 300,000 points. This difference emphasizes not only the improved confidence distribution but also the reduction in point count resulting from the selective filtering applied.

Figure 6 – Point Confidence histogram, (a) before filtering process, (b) after filtering



(a)

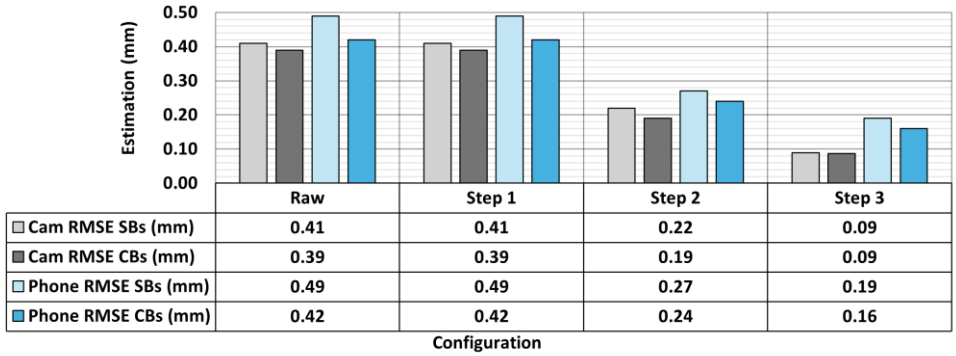


(b)

Source: Author (2024).

These results suggest that, while both image sets benefited from the filtering process, the camera set presented a higher density of points, possibly due to superior initial point quality and lower image noise levels. This substantial decrease in point density across both models reflects a careful elimination of redundant and low-accuracy points, contributing to more precise and reliable 3D model reconstructions, as presented in **Figure 7**.

Figure 7 – Assessment of the Quality of Point Filtering



Source: Author (2024).

For the 3D model derived from the camera image set (Cam), average Root Mean Square Error (RMSE) values improved from 0.41 mm in the unfiltered (Raw) configuration to 0.09 mm after filtering. The reconstruction from the smartphone image set (Smart) exhibited a less pronounced improvement, with RMSE decreasing from 0.49 mm and 0.42 mm for SBs and CBs in the Raw set to 0.19 mm and 0.16 mm, respectively, at the end of the filtering process.

The significant reduction in RMSE highlights how filtering can substantially enhance the accuracy and quality of 3D reconstructions, mitigating distortions and inaccuracies inherent in raw data sets, as supported by prior studies (e.g., Capolupo et al., 2020; Stark et al., 2022), especially when using consumer-grade imaging equipment.

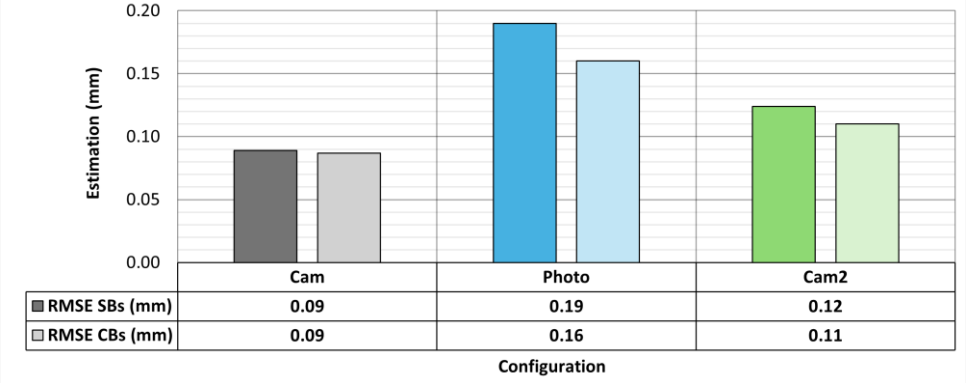
ASSESSMENT OF THE CONSISTENCY OF 3D MODELING PRODUCTS

Based on the results obtained from the filtering process, as previously discussed, the model generated from the camera image set was selected as the reference model for subsequent experiments. This decision was guided by the higher geometric precision and lower error values observed in the camera-based model compared to those generated from the smartphone image set, highlighting the camera's advantage in producing accurate spatial data for 3D modeling.

To assess the consistency of 3D models produced using consumer-grade imaging equipment, an additional image set was acquired with the camera positioned at a 45° tilt relative to the ground plane. This configuration aims to test the model's robustness under varying capture angles, which can introduce different challenges in alignment and scale accuracy.

Figure 8 presents the RMSE for SBs and CBs in this new 3D reconstruction, called Cam2, along with the previously discussed values for the Cam and Phone image sets.

Figure 8 – Assessment of the 3D Quality of Modeling of Set of Images



Source: Author (2024).

The RMSE obtained from the camera image sets, Cam and Cam2, showed a slight advantage over the values from the Phone set. Both Cam and Cam2 configurations demonstrated significant similarity to each other, with a marginally better performance observed in the Cam configuration.

After the complete processing of all image sets using the SfM technique, and given the absence of referenced control points, each model was initially referenced to an arbitrary coordinate system, defined according to the camera's internal geometry.

To enable a direct comparison between the generated models, it was necessary to align these systems to a common reference frame. Therefore, the coordinate system of the Cam model was chosen as the reference, and the remaining models were transformed into this system, allowing for consistent and accurate comparative analyses.

Table 4 the coordinate transformation residuals, categorized by axis (X, Y, Z) and control point distribution model (AT, IT, AIT, ECT), for both the Cam2 and Phone datasets relative to the reference coordinate system.

Table 4 – Residuals of the coordinate transformations for the Cam2 and Phone models relative to the Cam model

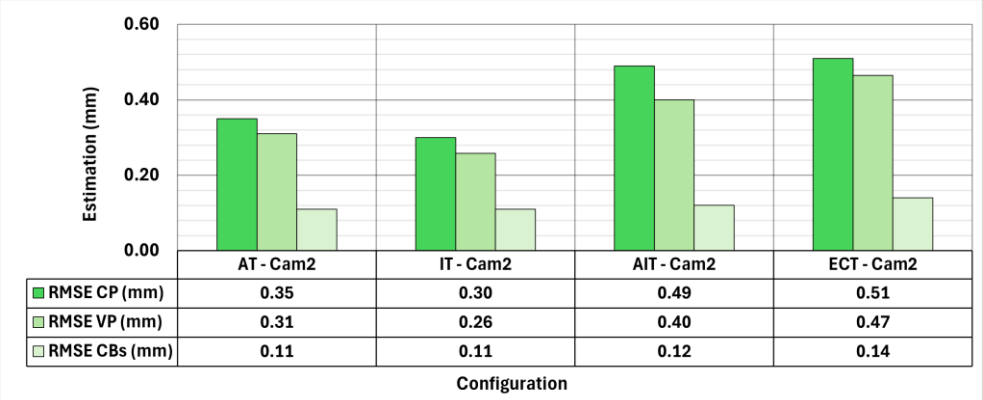
Cam2	AT	IT	AIT	ECT
X (mm)	0.25	0.18	0.28	0.32
Y (mm)	0.18	0.17	0.25	0.29
Z (mm)	0.10	0.08	0.14	0.19

Phone	AT	IT	AIT	ECT
X (mm)	0.25	0.17	0.33	0.39
Y (mm)	0.22	0.16	0.31	0.37
Z (mm)	0.12	0.09	0.18	0.26

The residuals demonstrate that the AT and IT configurations consistently yielded the lowest deviations. In contrast, the AIT and ECT configurations exhibited higher residuals, suggesting greater sensitivity to control point distribution.

Figure 9 presents the RMSE values for Control Points (CP) and Verify Points (VP)- also known Check Points, as well as for CBs. These values are presented for each configuration of distribution and quantity of reference points used in the transformation of the Cam2 model to the Cam system.

Figure 9 – Coordinate Transformation Assessment - Cam2 to Cam



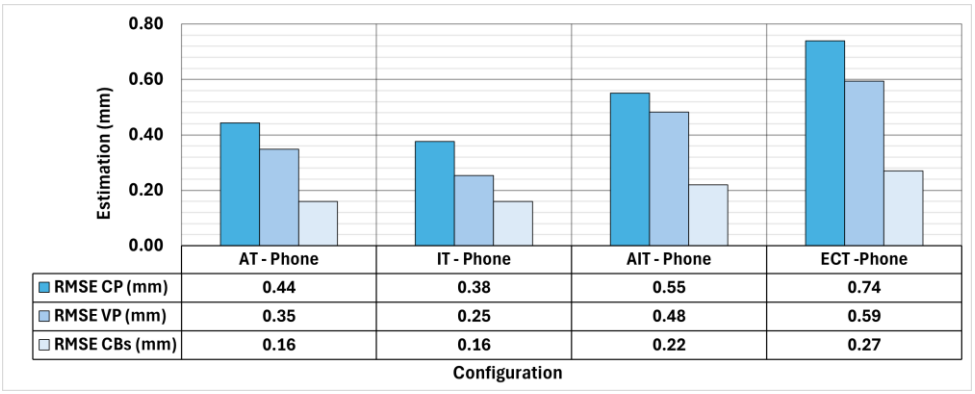
	AT - Cam2	IT - Cam2	AIT - Cam2	ECT - Cam2
RMSE CP (mm)	0.35	0.30	0.49	0.51
RMSE VP (mm)	0.31	0.26	0.40	0.47
RMSE CBs (mm)	0.11	0.11	0.12	0.14

Source: Author (2024).

The results indicate that the "AT" (Aligned Targets) and "IT" (Internal Targets) configurations preserved the RMSE for the CBs, matching the precision observed in the original Cam2 model before coordinate transformation, while other configurations showed a slight increase in RMSE, indicating a minor degradation in accuracy. For the CP and VP measurements, the "Internal Targets" configuration achieved the best results, with RMSE of 0.30 mm and 0.26 mm, respectively, underscoring its effectiveness in achieving accurate alignment.

Figure 10 displays the quality parameter RMSE for the coordinate transformation from Phone to the Cam model for each configuration and number of reference points.

Figure 10 – Coordinate Transformation Assessment - Phone to Cam



	AT - Phone	IT - Phone	AIT - Phone	ECT -Phone
RMSE CP (mm)	0.44	0.38	0.55	0.74
RMSE VP (mm)	0.35	0.25	0.48	0.59
RMSE CBs (mm)	0.16	0.16	0.22	0.27

Source: Author (2024).

In alignment with the results from the Cam2 transformation, the "AT" and "IT" configurations maintained RMSE for the CBs consistent with the untransformed results. In this case, the "IT" configuration again yielded the most accurate CP and VP values, with RMSEs of 0.38 mm and 0.25 mm, respectively.

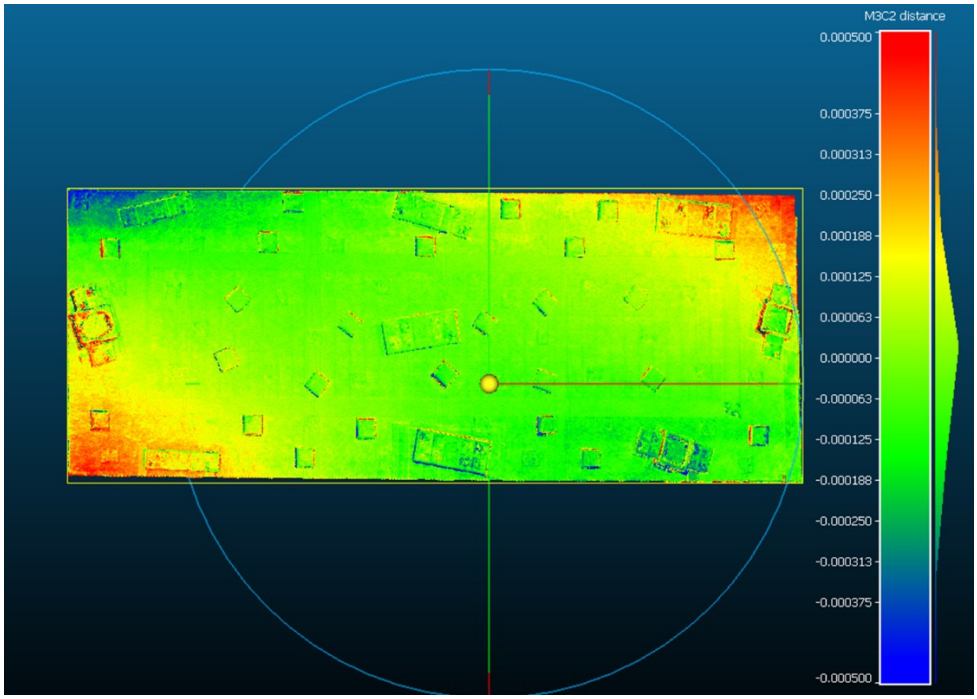
These results highlight the reliability of the "IT" configuration for maintaining model accuracy during coordinate transformation, particularly when transforming between image sets captured with different consumer-grade imaging equipment.

After the coordinate transformation, to reference all models from common control points, a comparative analysis was performed between the densified point

cloud of the reference set and the additional clouds generated using the SfM technique. This analysis, conducted in CloudCompare software, utilized the M3C2 method to evaluate geometric discrepancies between models, highlighting any structural variations and precision differences among the reconstructions, as discussed in Liao and Wood (2020).

Figure 11 displays the comparison between the 3D point clouds generated from the Cam and Cam2 image sets, emphasizing the geometric differences based on distances calculated through the M3C2 technique.

Figure 11 – Comparation between the 3D Dense Clouds – Cam and Cam2

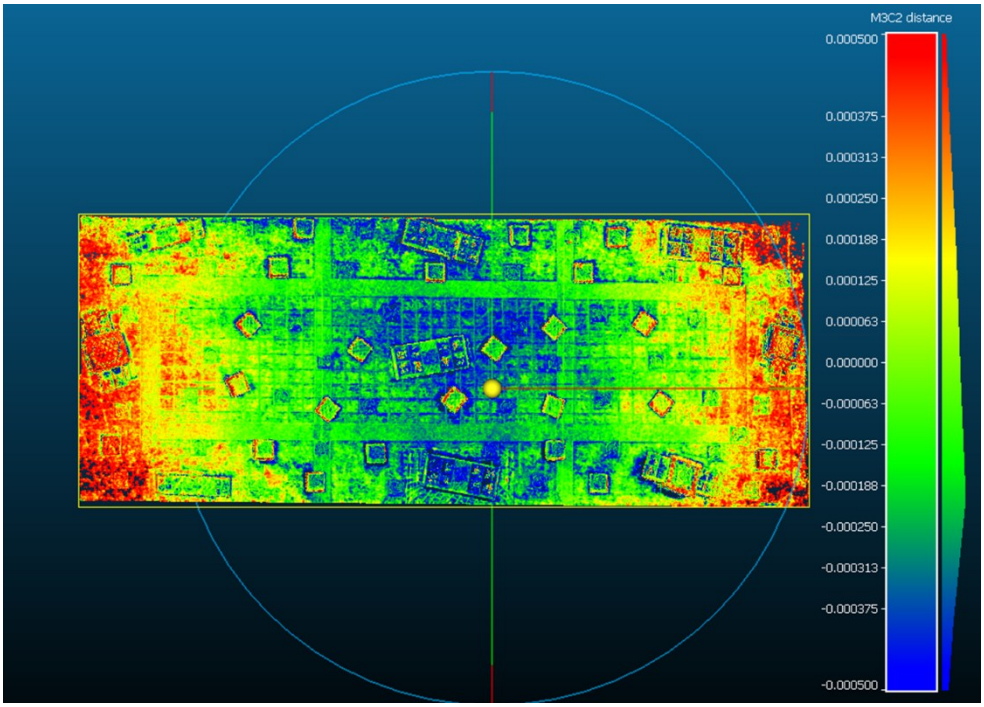


Source: Author (2024).

The Cam and Cam2 point clouds showed minimal variation between them, despite the altered camera angles used to capture the Cam2 set. This consistency suggests a high degree of robustness and quality in the camera-generated images, as similarly noted in previous assessments (Moraes; Silva, 2024). The most noticeable variations occurred at the model edges; however, within the primary area of interest — specifically, the internal region of the wooden board — discrepancies remained limited to around ±0.15 mm. These results reflect excellent precision and consistency in the 3D reconstruction of the central area, which is crucial for applications requiring high geometric fidelity, confirming that the image capture and filtering methods effectively preserved essential structural details.

Figure 12, in turn, presents the comparison between the 3D point clouds generated from the Cam and Phone image sets, with distances calculated via the M3C2 method.

Figure 12 – Comparation between the 3D Dense Clouds – Cam and Phone



Source: Author (2024).

Unlike the previous comparison, this analysis of the Cam and Phone sets revealed substantial differences throughout the area studied. This can be attributed to the lower image resolution of the smartphone, which lacks the detail of the camera-generated images. The impacts of these discrepancies were most apparent in regions with minimal surface texture variation, such as the plastic-coated surfaces of the scale elements and the cubes covered with silver tape, where noisy, low-precision zones led to distortions affecting surrounding model areas.

These outcomes highlight the critical role of capture equipment in determining the final accuracy of 3D models, particularly in areas with homogeneous textures where high-quality imaging is essential to avoid introducing noisy regions. **Figure 13** illustrates this effect in orthophotos generated by the SfM process for the Cam2 and Phone sets, focusing on a common region.

Figure 13 – Comparation of details levels in Orthophotos Cam2 and Smart

a b

Source: Author (2024).

In **Figure 13a**, derived from the Cam2 set, the edges of elements are clearly defined, allowing for precise visualization of detailed features. Conversely, **Figure 13b,** corresponding to the orthophoto from the Phone set, shows significant blurring in the same region, obscuring details and making it impossible to detect a comparable level of detail to that seen in **Figure 13a**.

These results indicate that while smartphones and similar consumer devices can achieve viable 3D reconstructions with quality in the order of millimeters or better, as pointed out in Boboc et al. (2019), specialized equipment, such as cameras with interchangeable lenses, remains essential in projects where high-fidelity outputs are crucial. This need is particularly pronounced in applications demanding detailed structural features and precise surface accuracy.

In conclusion, the choice of imaging equipment plays a decisive role in the fidelity of 3D reconstructions. High-quality cameras substantially enhance the accuracy and detail in critical areas, helping to mitigate issues such as noise and blurring that are more prevalent with lower-resolution devices. However, with the rapid advancements in smartphone imaging technology, the current advantage of high-end cameras may narrow shortly. Professionals will increasingly need to assess the specific demands of each project and the capabilities of available equipment to ensure compatible modeling.

LIMITATIONS OF THE STUDY

This study assessed the use of consumer-grade imaging equipment for 3D modeling in close-range laboratory environments for structural testing, highlighting key limitations of the experimental approach.

Coordinate Transformation: The lack of referenced control points required to coordinate conversions using common points across 3D models. This necessitated a broader photographic capture area to ensure an accurate representation of both the object and its surroundings. However, this approach presented challenges in laboratory spaces, where limited room and restricted viewpoints increase time demands.

Smartphone Usage: While modern smartphones have improved imaging capabilities, their current reliance on post-processing algorithms for noise reduction and detail enhancement negatively impacts 3D modeling accuracy. These algorithms had to be disabled, complicating the capture process and exposing limitations in brightness control and detail resolution due to smaller sensors. Despite these challenges, continued advancements in smartphone imaging technology may make them suitable for millimeter-level 3D modeling, particularly in contexts that do not demand submillimeter precision.

CONCLUSIONS

This study assesses the differences and quality levels achieved in 3D modeling using the SfM technique, comparing cameras with interchangeable lenses and smartphones as imaging devices. Through this comparative analysis, we sought to identify both the limitations and the potential of each device for producing

accurate three-dimensional models, focusing on geometric detail, positional accuracy, and the capacity for detailed surface representation of the specimen.

Photographic capture and SfM processing were conducted to generate high-quality 3D models, achieving precision at the millimeter scale or finer, suitable for structural analyses in controlled, short-distance laboratory environments. To maintain consistent lighting across the object, auxiliary lighting systems were utilized, minimizing light variability that could otherwise compromise capture consistency. Additionally, submillimeter-precision scale bars were incorporated to ensure accurate model scaling. The wooden test specimen was enhanced with artificial texture to improve reference point detection and enhance the SfM technique performance in generating a precise model.

Within this context, cameras with interchangeable lenses outperformed smartphones by offering greater control over light exposure and capturing higher-resolution details. This resulted in 3D models with RMSE around 0.1 mm and sharper details across detected elements. While smartphones also produced models with RMSE close to 0.2 mm, the level of detail was noticeably noisier and less sharp. This limitation may restrict smartphone suitability for reconstructing objects that require high levels of detail, especially in analyses that depend on precision and visual clarity.

The lower quality of smartphone images was also evident in surface representations, which were noisier and less defined, particularly in areas with uniform texture on the object's surface. Additionally, a subsequent model was developed through a new capture process with the camera, achieving quality and detail consistent with the initial model. This consistency indicates that the accuracy of the SfM technique is reliably repeatable when using cameras with interchangeable lenses, strengthening its validity for high-precision modeling.

Based on the experimental results, several key recommendations are proposed to ensure optimal outcomes when applying SfM in high-precision applications: (1) Select imaging equipment and configurations that yield a GSD compatible with the level of detail required for the intended analysis; (2) Use consistent and diffuse lighting conditions to avoid radiometric inconsistencies across the object; and (3) Carefully plan the distribution and quantity of control points, especially when the object exhibits fine geometric complexity, to avoid overfitting and ensure transformation accuracy.

In conclusion, while smartphones present a convenient and accessible option, their limitations in optical control and sensor quality make them less effective than cameras with interchangeable lenses for high-precision 3D modeling. Nevertheless, smartphones may be appropriate for preliminary surveys or for scenarios where submillimeter precision is not critical. Considering the rapid technological advancements in smartphone imaging, future studies should replicate this work to evaluate new developments in quality and performance across different imaging equipment. This continued evaluation will help define when smartphones can be reliably employed for specific tasks in structural monitoring, geotechnics, or archaeology, ultimately broadening their practical utility in these technical fields.

Filtragem de nuvem de pontos e desempenho de dispositivos em SfM: comparando câmeras e smartphones para modelagem 3D

RESUMO

A técnica Structure from Motion (SfM) tem sido amplamente empregada na modelagem 3D a partir de imagens digitais, destacando-se pela acessibilidade e compatibilidade com dispositivos convencionais, como câmeras e smartphones. No entanto, a qualidade dos modelos gerados depende diretamente da adequação do conjunto de imagens e de etapas de refinamento do processo, como a filtragem da nuvem de pontos, frequentemente subestimadas na literatura. Este estudo investiga a eficácia de diferentes dispositivos de captura na modelagem 3D por SfM, bem como o impacto da filtragem da nuvem de pontos na acurácia dos modelos gerados. Para isso, foram analisadas modelagens obtidas a partir de imagens capturadas por câmera e smartphone, com e sem a aplicação de filtragem na nuvem de pontos esparsa. Os resultados indicam que a filtragem é essência para a obtenção de uma alta qualidade dos modelos gerados por câmeras, resultando em um RMSE de 0,1 mm e maior detalhamento do objeto. Entretanto, os modelos obtidos com smartphones demonstraram potencial competitivo. Esses achados ressaltam a importância de estratégias de refinamento na modelagem SfM e contribuem para otimizar seu uso em diferentes contextos de captura.

PALAVRAS-CHAVE: Structure from Motion. Reconstruction Uncertainty. Projection Accuracy. Reprojection Error. Smartphone.

ACKNOWLEDGEMENT

The authors thank the São Carlos School of Engineering for all the support. This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 – process number: 88882.379118/2019-01.

REFERENCES

An, P., Fang, K., Jiang, Q., Zhang, H., Zhang, Y. (2021). Measurement of rock joint surfaces by using smartphone structure from motion (SFM) photogrammetry. **Sensors (Switzerland)**, *21*(3), 1–22. https://doi.org/10.3390/s21030922

Andrei, O. C. (2006). 3D affine coordinate transformations. **Geometria**, *3091*.

Barba, S., Barbarella, M., Di Benedetto, A., Fiani, M., Gujski, L., Limongiello, M. (2019). Accuracy assessment of 3d photogrammetric models from an unmanned aerial vehicle. **Drones**, *3*(4), 1–19. https://doi.org/10.3390/drones3040079

Boboc, R. G., Girbacia, F., Postelnicu, C. C., Girbacia, T. (2019). Evaluation of using mobile devices for 3D reconstruction of cultural heritage artifacts. VR Technologies in Cultural Heritage: First International Conference, VRTCH 2018, Brasov, Romania, May 29--30, 2018, Revised Selected Papers 1, 46–59.

Brown, D. C. (1971). Close-Range Camera Calibration. **Photogramm. Eng**, *8*(37), 855–866.

Caldera-Cordero, J. M., Polo, M.-E. (2019). Analysis of free image-based modelling systems applied to support topographic measurements. **Survey Review**, *51*(367), 300–309.

Capolupo, A. (2021). Accuracy assessment of cultural heritage models extracting 3D point cloud geometric features with RPAS SfM-MVS and TLS techniques. **Drones**, *5*(4). https://doi.org/10.3390/drones5040145

Capolupo, A., Saponaro, M., Mondino, E. B., & Tarantino, E. (2020). Combining interior orientation variables to predict the accuracy of rpas-sfm 3D models. **Remote Sensing**, *12*(17). https://doi.org/10.3390/RS12172674

Creus, P. K., Sanislav, I. V., Dirks, P. H. G. M. (2021). Application of SfM-MVS for mining geology: Capture set-up and automated processing using the Dugald River Zn-Pb-Ag mine as a case study. **Engineering Geology**, *293*(June), 106314. https://doi.org/10.1016/j.enggeo.2021.106314

Di Filippo, A., Antinozzi, S., Dell'amico, A., Sanseverino, A. (2022). a Statistical Analysis for the Assessment of Close-Range Photogrammetry Geometrical Features. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archive***s*, *48*(2/W2-2022), 31–38. https://doi.org/10.5194/isprs-archives-XLVIII-2-W2-2022-31-2022

Garcia, M. V. Y., Oliveira, H. C. de. (2021). The Influence pf Flight Configuration, Camera Calibration, and Ground Control Points for Digital Terrain Model and Orthomosaic Generation Using Unmanned Aerial Vehicles Imagery. **Boletim de Ciências Geodésicas**, *27*(2). https://doi.org/10.1590/s1982-21702021000200015

Hafeez, J., Jeon, H. J., Hamacher, A., Kwon, S. C., Lee, S. H. (2018). The effect of patterns on image-based modelling of texture-less objects. **Metrology and Measurement Systems**, *25*(4), 755–767. https://doi.org/10.24425/mms.2018.124883

James, M. R., Robson, S., d'Oleire-Oltmanns, S., Niethammer, U. (2017). Optimising UAV topographic surveys processed with structure-from-motion: Ground control quality, quantity and bundle adjustment. **Geomorphology**, *280*, 51–66. https://doi.org/10.1016/j.geomorph.2016.11.021

Jaud, M., Kervot, M., Delacourt, C., Bertin, S. (2019). Potential of smartphone SfM photogrammetry to measure coastal morphodynamics. **Remote Sensing**, *11*(19). https://doi.org/10.3390/rs11192242

Leon, J. X., Roelfsema, C. M., Saunders, M. I., Phinn, S. R. (2015). Measuring coral reef terrain roughness using "Structure-from-Motion" close-range photogrammetry. **Geomorphology**, *242*, 21–28. https://doi.org/10.1016/j.geomorph.2015.01.030

Liao, Y., Wood, R. L. (2020). Discrete and distributed error assessment of UAS-SFM point clouds of roadways. **Infrastructures**, *5*(10), 1–31. https://doi.org/10.3390/infrastructures5100087

Liu, Y.-F., Cho, S., Spencer, B. F., Fan, J.-S. (2016). Concrete Crack Assessment Using Digital Image Processing and 3D Scene Reconstruction. **Journal of Computing in Civil Engineering**, *30*(1), 04014124. https://doi.org/10.1061/(asce)cp.1943-5487.0000446

Luhmann, T., Robson, S., Kyle, S., Boehm, J. (2023). **Close-Range Photogrammetry and 3D Imaging**: Vol. 4th Edition (4th editon). DE GRUYTER.

Micheletti, N., Chandler, J. H., Lane, S. N. (2015). Investigating the geomorphological potential of freely available and accessible structure-from-motion photogrammetry using a smartphone. **Earth Surface Processes and Landforms**, *40*(4), 473–486. https://doi.org/10.1002/esp.3648

Moraes, F. R. de, Silva, I. da. (2024). Assessment of Submillimeter Precision via Structure from Motion Technique in Close-Range Capture Environments. 1–23.

Nota, E. W., Nijland, W., de Haas, T. (2022). Improving UAV-SfM time-series accuracy by co-alignment and contributions of ground control or RTK positioning. **International Journal of Applied Earth Observation and Geoinformation**, *109*(March), 102772. https://doi.org/10.1016/j.jag.2022.102772

Over, J.-S. R., Ritchie, A. C., Kranenburg, C. J., Brown, J. A., Buscombe, D. D., Noble, T., Sherwood, C. R., Warrick, J. A., Wernette, P. A. (2021). Processing coastal imagery with Agisoft Metashape Professional Edition, version 1.6—Structure from motion workflow documentation.

Remondino, F. ;, Fraser, C., Remondino, F. (2006). Digital camera calibration methods: considerations and comparisons. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, *36*(5), 266–272. https://doi.org/10.3929/ethz-b-000158067

Stark, M., Rom, J., Haas, F., Piermattei, L., Fleischer, F., Altmann, M., Becht, M. (2022). Long-term assessment of terrain changes and calculation of erosion rates in an alpine catchment based on SfM-MVS processing of historical aerial images. How camera information and processing strategy affect quantitative analysis. **Journal of Geomorphology**, *July*. https://doi.org/10.1127/jgeomorphology/2022/0755

Starrett Company. (2024). Starrett Electronic Precision Hand Tools.

Tavani, S., Pignalosa, A., Corradetti, A., Mercuri, M., Smeraglia, L., Riccardi, U., Seers, T., Pavlis, T., Billi, A. (2020). Photogrammetric 3d model via smartphone gnss sensor: Workflow, error estimate, and best practices. **Remote Sensing**, *12*(21), 1–19. https://doi.org/10.3390/rs12213616

Tinkham, W. T., Swayze, N. C. (2021). Influence of agisoft metashape parameters on uas structure from motion individual tree detection from canopy height models. **Forests**, *12*(2), 1–14. https://doi.org/10.3390/f12020250

Triggs, B., Mclauchlan, P., Hartley, R., Fitzgibbon, A., McLauchlan, P. (2000). International Workshop on Vision Algorithms. 298–372. https://doi.org/10.1007/3-540-44480-7_21ï

Verma, A. K., Bourke, M. C. (2019). A method based on structure-from-motion photogrammetry to generate sub-millimetre-resolution digital elevation models for investigating rock breakdown features. **Earth Surface Dynamics**, *7*(1), 45–66. https://doi.org/10.5194/esurf-7-45-2019

Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., Reynolds, J. M. (2012). "Structure-from-Motion" photogrammetry: A low-cost, effective tool for geoscience applications. **Geomorphology**, *179*, 300–314. https://doi.org/10.1016/j.geomorph.2012.08.021

Woodget, A. S., Austrums, R. (2017). Subaerial gravel size measurement using topographic data derived from a UAV-SfM approach. **Earth Surface Processes and Landforms**, *42*(9), 1434–1443. https://doi.org/10.1002/esp.4139

Recebido: 13 fev. 2025
Aprovado: 06 jun. 2025
DOI: 10.3895/rbgeo.v13n2.19909
Como citar: MORAES, F. R.; SILVA, I. Point cloud filtering and device performance in SfM: comparing cameras and smartphones for 3D modeling. **R. bras. Geom.**, Curitiba, v. 13, n. 1, p. 168-191, jan./jun. 2025. Disponível em: <https://periodicos.utfpr.edu.br/rbgeo>. Acesso em: XXX.
Correspondência:
Francisco Roza de Moraes
Av. Trabalhador são-carlense, 400, CEP 13566-590, São Carlos, São Paulo, Brasil.